

# Modeling Atrazine Occurrence Patterns with Nonparametric Neural Networks

Michael Messner

USEPA Office of Ground Water and Drinking Water Standards and Risk Management Division



## ABSTRACT

OW's Atrazine Team wants to understand how well the Atrazine Rule performs with respect to short-term peaks in exposure, which some research indicates may be an important human health consideration. To understand the performance of the Atrazine Rule, the Team needed to assess the likelihood of compliance and noncompliance, given various occurrence levels and patterns. This poster describes how new occurrence data (provided through the Office of Pesticides Program's re-registration of the pesticide) were used to characterize a range of occurrence patterns.

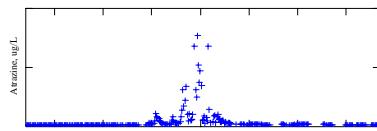
Re-registration data sets having low, moderate, and high variability were identified, and their occurrence patterns were modeled using nonparametric neural network methods. Although nonparametric in the statistical sense, the models contained numerous parameters, including one to describe serial correlation. One location's pattern required 15 parameters! Maximum likelihood parameter values were identified using greedy search algorithms.

The resulting occurrence patterns can be scaled to represent low, moderate, and high variability waters with various occurrence levels.

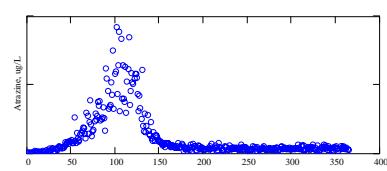
## OCCURRENCE PATTERNS

Data from Syngenta (registrar of the herbicide atrazine) were provided for dozens of source waters known or believed to be vulnerable to atrazine runoff from farmland. These data were reviewed to identify waters with different levels of seasonal variability. All of these waters were found to have less variability than seen in a historical data set provided by Missouri American Water Company. The Missouri American data are taken to represent a high level of seasonal variability. Three waters from Syngenta's 2003 re-registration set were taken to represent moderate, low, and zero seasonal variability patterns.

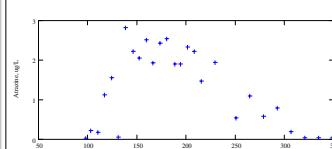
### High Seasonal Variability:



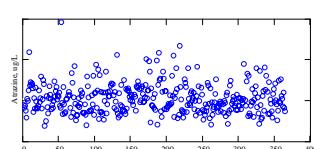
### Moderate Seasonal Variability:



### Low Seasonal Variability:



### Zero Seasonal Variability:



## NEURAL NET MODELS

Artificial Neural Network models of the following form were used to fit the moderate and low-variability occurrence patterns:

$$\ln(C_{\text{day}}) = \rho_0 + \frac{\rho_1}{1 + e^{-\rho_3 - \rho_4 \cdot \text{Day}}} + \frac{\rho_2}{(1 + e^{-\rho_5 - \rho_6 \cdot \text{Day}})}$$

The high-variability water data set, with its daily monitoring, exhibited serial correlation that was not evident in the re-registration waters' weekly or bi-weekly monitoring. It also seemed to include two strong occurrence peaks. Additional terms were added to the model to explain the second peak and a parameter was added to explain the autoregressive behavior:

$$\ln(C_{\text{day}}) = \rho_0 + \frac{\rho_1}{1 + e^{-\rho_3 - \rho_6 \cdot \text{Day}}} + \frac{\rho_2}{(1 + e^{-\rho_7 - \rho_8 \cdot \text{Day}})} + \frac{\rho_3}{(1 + e^{-\rho_9 - \rho_{10} \cdot \text{Day}})} + \frac{\rho_4}{(1 + e^{-\rho_{11} - \rho_{12} \cdot \text{Day}})}$$

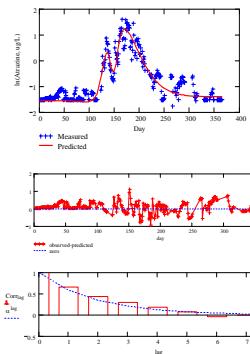
## ESTIMATION

The high-variability pattern is described below, as it is the most complex (interesting).

A greedy search algorithm was used to find maximum likelihood parameters ( $\rho$ ). When the algorithm began to slow down (because it was near optimum), a solver was used to find where the partial derivatives were zero. Next, residuals were plotted and studied. A correlogram suggested autoregressive behavior and a maximum likelihood autoregression coefficient ( $\alpha$ ) was found.

### Parameter Estimate

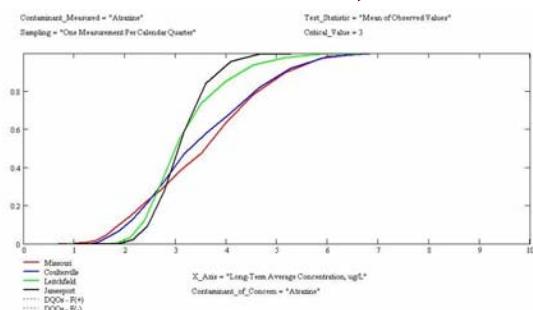
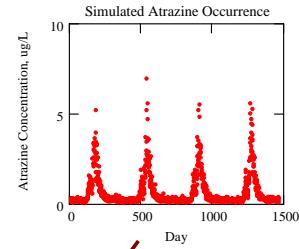
Parameter	Estimate
$\rho_0$	-10.185
$\rho_1$	4.810
$\rho_2$	4.549
$\rho_3$	3.964
$\rho_4$	4.091
$\rho_5$	-18.57
$\rho_6$	0.142
$\rho_7$	29.082
$\rho_8$	-0.206
$\rho_9$	.27.772
$\rho_{10}$	0.180
$\rho_{11}$	7.819
$\rho_{12}$	-0.041
$\alpha$	0.633
$\sigma$	0.279



**Figures to the left show**  
a. nice model fit (observed versus predicted),  
b. strong evidence of serial correlation in the residuals plot, and  
c. correlogram showing autoregressive behavior with parameter  $\alpha = 0.633$ .

## SIMULATIONS and RESULTS

With the maximum likelihood parameters, it is a simple matter to simulate additional years' data under the same occurrence pattern. To simulate quarterly monitoring, one only needs to grab one day's concentration from each calendar quarter. The problem is that this portrays only one, rather low-occurrence scenario. Parameter  $\rho_0$  can be adjusted to, in essence, scale the occurrence pattern by a multiplicative factor. The resulting pattern has the same shape, but different long-term average and different number of days per year above selected thresholds. By controlling parameter  $\rho_0$ , performance of the candidate strategies could be studied over a range of occurrence levels.



The results of these modeling efforts are numerous graphs like the one above, showing probabilities of "taking action" as a result of employing the Atrazine Rule and other strategies.

This remains a work in progress. The author welcomes comments & feedback. Please feel free to contact by phone (202-564-5268) or E-mail (messner.michael@epa.gov).

## REFERENCES

- Andrew Merritt, Syngenta Crop Protection, Inc., "2003 Atrazine Monitoring Report", Jan. 2004.
- Herbert H.K. Lee, *Bayesian Nonparametrics via Neural Networks*, ASA/SIAM, 2004.

**DISCLAIMER:** The findings and conclusions in this poster are those of the author and do not necessarily represent the views of the Environmental Protection Agency.



**epascience**forum  
Collaborative Science  
for Environmental Solutions



2005  
epascienceforum.org